Brennan, R.L. (2004). *Some perspectives on inconsistencies among measurement models* (Research Report No. 8). Iowa City, IA: University of Iowa, Center for Advanced Studies in Measurement and Assessment.

1. **What constitutes a replication?**

Generalizability theory requires specification of what facets of the measurement procedure are fixed or random over replications. This typically includes items, occasions, tasks, and raters.

1. What are the intended (possibly idealized) replications of the measurement procedure?
2. What are the characteristics of the data actually available or to be collected to estimate reliability?

When a facet is intended to be random (a) but is essentially fixed (b) in a particular data set, the reliability coefficient is overestimated since error variance will be underestimated.

In IRT, there is no direct sampling model of content – no sampling of items from a domain. The item parameters are based on a fixed set of items on the test. So the SEMs in IRT will be smaller than those in CTT or G-Theory.

1. **What are True Scores?**

In CTT and G-Theory, the models are based on expected-value based notions of true scores; whereas in IRT, the true score is a known-to-be-true attribute of the object of measurement.

The 3PL model approaches an expected-value notion of true scores, since it recognizes that low ability test takers have a non-zero probability of correct response suggesting that they don’t “know” the answer, in a more platonic sense.

In 1PL and 2PL models, low ability examinees have near zero probability of correct response. In addition, items are fixed in IRT – true scores based on the TCC are for a fixed set of items.

In CTT, true score is defined as expected value of observed scores over forms that are similar; in G-Theory universe scores are expected values over randomly parallel forms.

1. **What is Error?**

Depends on the definition of true score and estimation method.

Also see:

Brennan, R.L. (1995). The conventional wisdom about group mean scores. *Journal of Educational Measurement, 14*, 385–396.

Brennan, R.L. (2001). Some problems, pitfalls, and paradoxes in educational measurement. *Educational Measurement: Issues and Practice, 20* (4), 6-18.